CRC

METHODS *in* NUCLEIC ACIDS RESEARCH

Jim D. Karam
Lee Chao
Gregory W. Warr

CRC PRESS

Methods in Nucleic Acids Research

Editors

Jim D. Karam, Ph.D.

Lee Chao, Ph.D.

Gregory W. Warr, Ph.D.

Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Charleston, South Carolina

CRC Press
Boca Raton Ann Arbor Boston

Library of Congress Cataloging-in-Publication Data

Methods in nucleic acids research/editors, Jim D. Karam, Lee Chao,
Gregory W. Warr.
p. cm.
Includes bibliographical references.
Includes index.
ISBN 0-8493-5311-4
1. DNA--Research--Technique. 2. Nucleic acids--Research-
-Technique. 3. Gene expression--Research--Technique. I. Karam,
Jim D. II. Chao, Lee. III. Warr, Gregory W.
[DNLM: 1. DNA. 2. Research--methods. 3. RNA. QU 58 M5922]
QP624.M49 1990
574.87'3282--dc20
DNLM/DLC
for Library of Congress 90-2431
CIP

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida, 33431.

International Standard Book Number 0-8493-5311-4

Library of Congress Card Number 90-2431
Printed in the United States

PREFACE

We embarked on this project realizing that there are, on the market, several excellent laboratory manuals in molecular biology, and yet feeling that we and our students needed something which paid more attention to principles and also provided background information on the evolution of new technology in this field. The task was made difficult by the incredibly rapid rate of technical progress, which constantly nagged at us and our authors as we tried to complete this manual in a timely manner. We hope that we have been successful in keeping only one or two steps behind progress and that this book will prove to be a valuable addition to libraries of our colleagues and their students. We have tried to select experts in newly emerging techniques as well as in areas (many of them technically demanding) that are not frequently covered in other methods books. As editors, we also tried to fill in some gaps and details that we thought would add balance, i.e., the short chapters on expression plasmids in prokaryotes (Karam) and rapid immunoscreening of expression libraries (Chao et al.). About midway through the process of receiving and editing manuscripts for the book, polymerase chain-reaction (PCR) technology started mushrooming at a pace that we had not anticipated. Considering that there are several good sources of information on this technology and its application to different situations (substantial methods books in their own right), we decided to invite a chapter on an application of PCR that illustrated the basic principle and that most of us could use right away, hence the short chapter on direct amplification of DNA from phage plaques (Miller and Jozwik). Deshler et al. also include a section on the principles and practice of PCR utilizing DNA or RNA as the starting material, and Budowle et al. deal with the use of PCR in forensic applications. Some techniques, although very important, are so well covered elsewhere and such good kits are sold for their execution (e.g., cDNA library construction) that we decided to omit them from this volume. On the other hand, some methods have become so central and important for most investigations in molecular genetics that we wanted to include a good selection of these: genomic DNA library construction (Hinds et al.), electrophoretic separation of large DNA molecules (pulsed-field gel electrophoresis and related methods, Bentley et al.), the use of prokaryotic (Karam) and eukaryotic (Maruyama et al.) expression vectors, and DNA sequencing methods (Slatko).

Many of the chapters include experimental details pertaining to widely used methods of DNA preparation, quantitation, and analysis. In terms of basic approaches to the analysis of gene structure and function, this volume incorporates a survey of approaches that, in addition to those mentioned above, includes the preparation of synthetic DNA for use in a variety of applications (Deshler et al.), *in vitro* site-directed mutagenesis (Menick), and *in vitro* methods for the analysis of eukaryotic transcription (Ruteshouser et al.) and translation (Wahba and Dholakia). A significant portion of the volume is directed to methods for the detection of protein/nucleic acid interactions, including gel-retardation assays (Revzin), DNA footprinting for the analysis of sequence specific proteins (Pauli et al.), and RNA footprinting techniques, a relatively new tool for the analysis of RNA function (McPheeters). We have also included approaches for studies in newly emerging areas of interest in molecular biology, e.g., the chapters on human genome mapping (Lathrop and Nakamura), the use of molecular biology in forensic analysis (Budowle et al.), the detection of mRNA sequences by *in situ* hybridization (Awgulewitsch and Utset), and the production and use of antibodies to synthetic oligopeptides (Schluter).

In all the invited chapters we asked the authors to provide brief surveys of the literature that would allow the reader to explore further not only the principles of the techniques, but also other experimental details. We think that historical or experimental perspectives on how new technology has developed are of importance in the educational process for the next

generation of biological scientists and hope that in some measure this volume can be of both practical bench-top utility and educational value.

Jim D. Karam, Ph.D.
Lee Chao, Ph.D.
Gregory W. Warr, Ph.D.

THE EDITORS

Jim D. Karam, Ph.D., is Professor of Biochemistry and Molecular Biology at the Medical University of South Carolina in Charleston. He obtained his undergraduate training at the American University of Beirut in Lebanon, receiving his B.Sc. degree in 1958. He earned the Ph.D. degree in Biochemistry in 1965 from the University of North Carolina at Chapel Hill. As a recipient of a Postdoctoral Fellowship from the National Institute of General Medical Sciences from 1965 to 1967, Dr. Karam carried out studies at the Cold Spring Harbor Laboratory in New York on the role of DNA polymerase in control of mutation rates. He continued these studies as a Research Assistant Professor at the University of Connecticut, Storrs from 1967 to 1968. Before joining the faculty at the Medical University of South Carolina in 1971, Dr. Karam was Research Associate (1968 to 1970) and Associate (1970 to 1971) at the Sloan-Kettering Institute for Cancer Research in New York City.

Dr. Karam is a member of the American Society of Biochemistry and Molecular Biology, the Genetics Society of America, and the American Society for Microbiology. He has served on several grant review panels for the National Institutes of Health and National Science Foundation. His research interests are in the area of control of gene expression and he has published papers dealing with human, *Drosophila*, and microbial genetic systems. From 1974 to 1979 he was recipient of a Research Career Development Award from the National Institute of General Medical Sciences. He has also been the recipient of many research grants during his career. Currently, Dr. Karam's investigations into the genetic control of phage T4 DNA replication and the roles of translational repressors in control of protein biosynthesis are being supported by grants from the National Institutes of Health and the National Science Foundation.

Lee Chao, Ph.D., is Professor of Biochemistry and Molecular Biology at the Medical University of South Carolina, Charleston, South Carolina. He graduated from the National Taiwan University in 1963 and received his Masters degree in 1966 from Utah State University. In 1970 he completed his Ph.D. training at Iowa State University. Beginning in 1970 he moved to the Molecular Biology Laboratory at the University of Connecticut for his postdoctoral training. He moved to the Medical University of South Carolina in 1974 as an Assistant Professor and was appointed Professor in 1985. He is a member of the American Society of Biochemistry and Molecular Biology, the American Society of Hypertension, the American Heart Association, and the American Association for the Advancement of Science. Dr. Chao has been the recipient of research awards from the National Institutes of Health and has served as reviewer for the NIH and as a panelist for the National Research Council.

Dr. Chao is the author of over 50 research articles, reviews, and book chapters. He has presented over 30 invited lectures at national and international meetings and at universities and institutes. His current major research interests include the molecular mechanisms of multigene family regulation and evolution and the genetic basis of hypertensive diseases.

Gregory W. Warr, Ph.D., is Professor of Biochemistry and Molecular Biology at the Medical University of South Carolina, Charleston, South Carolina. He graduated from the University of Cambridge in 1970 and received his Ph.D. degree from the University of London in 1973. After postdoctoral work at the University of Edinburgh and the Walter and Eliza Hall Institute in Melbourne, Australia, Dr. Warr was a staff scientist at the National Cancer Institute Cancer Research Center in Frederick, Maryland. He moved to the Medical University of South Carolina in 1980 as an Associate Professor and was appointed Professor

in 1986. Dr. Warr was a recipient of a Senior International Fellowship of the Fogarty International Center (1987 to 1988) for study at the University of Cambridge.

Dr. Warr is a member of the American Association of Immunologists, the American Society of Zoologists, the International Society for Developmental and Comparative Immunology, the Society for Invertebrate Pathology, and the British Society for Immunology. In 1991 and 1992 he will be Chairman of the Division of Comparative Immunology of the American Society of Zoologists. He has served as a member of the editorial board of the *Journal of Invertebrate Pathology* (1983 to 1987) and is currently an Associate Editor of *Developmental and Comparative Immunology*.

Dr. Warr has been the recipient of research awards from the National Science Foundation and the National Institutes of Health and has served as a reviewer for these two agencies. Dr. Warr is the author of over 100 research articles, reviews, and book chapters and has co-edited 3 books. His current research interests center on the evolution of the immunoglobulin molecule and its genes.

CONTRIBUTORS

Dwight E. Adams, Ph.D.
DNA Analysis Unit
Federal Bureau of Investigation
Washington, D.C.

Robert C. Allen, Ph.D.
Department of Pathology and Laboratory Medicine
Medical University of South Carolina
Charleston, South Carolina

Naoko Arai, M.D., Ph.D.
Department of Molecular Biology
DNAX Research Institute
Palo Alto, California
and National Institute of Health
Tokyo, Japan

Alexander Awgulewitsch, Ph.D.
Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Charleston, South Carolina

Kevin L. Bentley, Ph.D.
Department of Biology
Yale University
New Haven, Connecticut

Charles J. Bieberich, Ph.D.
Laboratory of Virology
American Red Cross
Rockville, Maryland

Bruce Budowle, Ph.D.
Forensic Science Research and Training Center
Federal Bureau of Investigation Academy
Quantico, Virginia

Julie Chao, Ph.D.
Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Charleston, South Carolina

Lee Chao, Ph.D.
Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Charleston, South Carolina

Steven Chao
Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Charleston, South Carolina

Benoit de Crombrugghe, M.D.
Department of Molecular Genetics
The University of Texas M. D. Anderson Cancer Center
Houston, Texas

James O. Deshler, Ph.D.
Department of Molecular Genetics
Beckman Research Institute
City of Hope
Duarte, California

Jaydev N. Dholakia, Ph.D.
Department of Biochemistry
The University of Mississippi Medical Center
Jackson, Mississippi

Kristin R. Hinds, Ph.D.
Department of Pediatrics
University of South Florida
All Children's Hospital
St. Petersburg, Florida

Catherine E. Jozwik, B.S.
Department of Microbiology
North Carolina State University
Raleigh, North Carolina

Bruce E. Kaplan, Ph.D.
Department of Molecular Genetics
Beckman Research Institute
City of Hope
Duarte, California

Jim D. Karam, Ph.D.
Department of Biochemistry and
Molecular Biology
Medical University of South Carolina
Charleston, South Carolina

Garrett P. Larson, M.A.
Department of Molecular Genetics
Beckman Research Institute
City of Hope
Duarte, California

G. Mark Lathrop, Ph.D.
Centre d'Etude du Polymorphisme
Humain
Paris, France

Gary W. Litman, Ph.D.
Department of Pediatrics
University of South Florida
All Children's Hospital
St. Petersburg, Florida

Sankar N. Maity, Ph.D.
Department of Molecular Genetics
The University of Texas M. D. Anderson
Cancer Center
Houston, Texas

Kazuo Maruyama, Ph.D.
Department of Molecular Biology
DNAX Research Institute
Palo Alto, California
and The Institute of Medical Science
University of Tokyo
Tokyo, Japan

David S. McPheeters, Ph.D.
Division of Biology
California Institute of Technology
Pasadena, California

Donald R. Menick, Ph.D.
Cardiology Division
Medical University of South Carolina
Charleston, South Carolina

Eric S. Miller, Ph.D.
Department of Microbiology
North Carolina State University
Raleigh, North Carolina

Atsushi Miyajima, Ph.D.
Department of Molecular Biology
DNAX Research Institute
Palo Alto, California
and The Institute of Medical Science
University of Tokyo
Tokyo, Japan

Yasuke Nakamura, M.D., Ph.D.
Division of Biochemistry
Cancer Institute
Tokyo, Japan

Urs Pauli, Ph.D.
Institute of Veterinary Virology
University of Bern
Bern, Switzerland

Arnold Revzin, Ph.D.
Department of Biochemistry
Michigan State University
East Lansing, Michigan

John J. Rossi, Ph.D.
Department of Molecular Genetics
Beckman Research Institute
City of Hope
Duarte, California

Frank H. Ruddle, Ph.D.
Department of Biology
Yale University
New Haven, Connecticut

E. Cristy Ruteshouser, Ph.D.
Department of Molecular Genetics
The University of Texas M. D. Anderson
Cancer Center
Houston, Texas

Samuel F. Schluter, Ph.D.
Department of Microbiology and
Immunology
Health Sciences Center
University of Arizona
Tucson, Arizona

Michael J. Shamblott, B.S.
Center of Marine Biotechnology
University of Maryland
Baltimore, Maryland

Barton E. Slatko, Ph.D.
New England Biolabs, Inc.
Beverly, Massachusetts

Gary S. Stein, Ph.D.
Department of Cell Biology
University of Massachusetts Medical Center
Worcester, Massachusetts

Janet L. Stein, Ph.D.
Department of Cell Biology
University of Massachusetts Medical Center
Worcester, Massachusetts

Yutaka Takebe, M.D., Ph.D.
Department of Molecular Biology
DNAX Research Institute
Palo Alto, California
and National Institute of Health
Tokyo, Japan

Manuel F. Utset, M.D., Ph.D.
Department of Pathology
Johns Hopkins Hospital
Baltimore, Maryland

André van Wijnen, M.S.
Department of Cell Biology
University of Massachusetts Medical Center
Worcester, Massachusetts

Albert J. Wahba, Ph.D.
Department of Biochemistry
The University of Mississippi Medical Center
Jackson, Mississippi

Huey-Mei Wang, Ph.D.
Department of Molecular Biology
DNAX Research Institute
Palo Alto, California
and The Institute of Medical Science
University of Tokyo
Tokyo, Japan

Gregory W. Warr, Ph.D.
Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Charleston, South Carolina

Kenneth L. Wright, Ph.D.
Department of Cell Biology
University of Massachusetts Medical Center
Worcester, Massachusetts

TABLE OF CONTENTS

I. PREPARATIVE METHODS

Chapter 1

PURIFICATION OF CHROMOSOMAL DNA FOR LIBRARY CONSTRUCTION

K. Hinds, M. J. Shamblott, and G. W. Litman

TABLE OF CONTENTS

I. INTRODUCTION

The construction of recombinant genomic DNA libraries is a critical step in the analysis of the genetic makeup of an organism. Early libraries were constructed in plasmid vectors.[1] While this approach was useful for the study of simple prokaryotic organisms, the size of more complex genomes, typically 3×10^9 bp (base pairs) or greater per haploid genome, requires the use of vectors capable of replicating large pieces of insert DNA. The vectors developed to facilitate the cloning of large fragments of DNA are specially engineered λ-phage[2] and constructs known as cosmids that consist of an antibiotic-resistant plasmid into which cos sites (cohesive ends) from λ have been cloned to allow *in vitro* packaging.[3] Many special features have been added to both types of vectors that enhance the analysis of the insert DNA. Certain vectors have been engineered to facilitate deriving hybridization probes for use in further experiments.

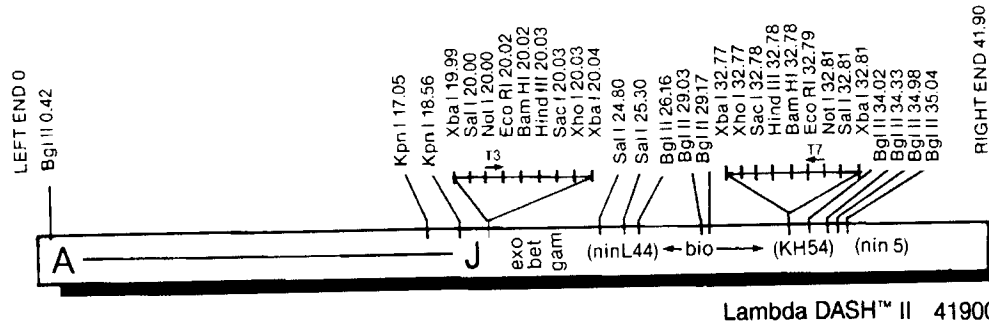
The λ vectors available for genomic library construction are known as replacement vectors. In these vectors, the central portion of the phage genome is not needed for lytic replication and can be replaced by genomic DNA from the animal of interest. Vectors are available that incorporate insert DNA of a variety of sizes up to 23 kb, and most currently available vectors have been designed to maximize the size of the insert DNA. The size constraints are provided by the *in vitro* packaging reaction. This reaction packages DNA (vector + insert) into empty phage protein coats when the DNA is between 78 and 105% of the size of wild-type λ and is flanked by two cos sites.[4] Most available vectors also utilize a selective system (based on the spi^+ phenotype) that does not allow growth of nonrecombinant bacteriophage. Recently developed vectors typically are engineered (by double digests of closely linked restriction sites) to prevent religation of the nonessential central fragment, eliminating the need for physical separation. In addition, these vectors contain the promoters for T3 and T7 RNA polymerases that facilitate rapid synthesis of insert-specific probes for use in chromosome walking and restriction mapping. Examples of λ vectors with these properties are Lambda DASH™ II (Stratagene), as illustrated in Figure 1A, and λGEM-11 (Promega).

Similar features have been engineered into cosmid vectors. One such vector system, pWE15 and 16,[5] available from Stratagene, is used as an example (Figure 1B). pWE15 has T3 and T7 promoters and a *Bam*HI cloning site flanked by *Not*I sites that permit excision of the intact genomic insert. The small size, ~8.8 kb, of the vector permits cloning of 30- to 42-kb genomic fragments, and the antibiotic resistance genes allow cloning in both prokaryotic and eukaryotic host systems.

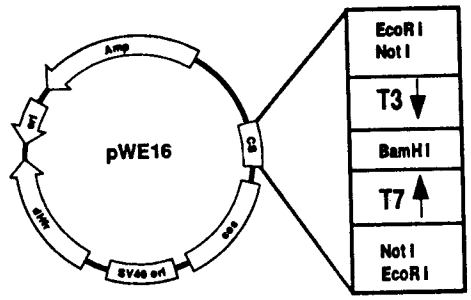
Since vectors capable of replicating large segments of insert DNA are available, the quality, particularly the size, of the genomic DNA to be cloned is of critical importance. Most cloning methods rely on restriction enzymes to generate clonable DNA. Any DNA ends which are caused by breaks, rather than being generated by enzymes, will compromise the cloning process. In order to realize the full potential of the phage and cosmid cloning systems, great care must be exercised in preparing purified high molecular weight genomic DNA for cloning.

Preparing genomic DNA for partial restriction-enzyme digestion requires special consideration. As mentioned above, the DNA must be prepared carefully to insure that it is large enough to yield approximately 20-kb and 30- to 40-kb fragments with two cohesive ends for λ and cosmid cloning, respectively. The cleavage should be random, although this is an ideal situation that cannot be achieved due to the nonuniform distribution of restriction sites in DNA. To enhance the generation of appropriate DNA fragments, the four base-pair cutting enzyme *Sau*3A is used. Partial enzyme cutting is achieved by enzyme dilution or time-course control.

The size of the library (the number of recombinant phage) needed to find a particular



A



B

FIGURE 1. Cloning vector maps. (A) The Lambda DASH® II vector showing the restriction enzyme sites used in arm preparation and cloning as well as the location of the promoters for the T3 and T7 RNA polymerases; (B) the pWE16 cosmid vector showing the *Bam*HI cloning site flanked by the T3 and T7 RNA polymerase promoters. (Maps courtesy of John C. Bauer, Stratagene, La Jolla, CA.)

DNA sequence is dependent on the size of the genome of the organism and the size of the insert DNA in each phage. If DNA sequence representation is completely random, the chance of finding a sequence of interest can be calculated using the formula.[1]

$$N = \frac{\ln (1 - P)}{\ln (1 - X/Y)}$$

where X is the insert size and Y is the size of the haploid genome. For a P (probability) of 0.99, it would be necessary to screen 9.2×10^2 λ recombinants (20-kb inserts) or 4.3×10^2 cosmid recombinants (45-kb inserts) if the genome size is 4.2×10^6 bp (*Escherichia coil*). For human DNA (3.3×10^9-bp haploid genome), a P of 0.99 would require 7.6×10^5 λ-recombinants or 3.4×10^5 cosmid recombinants. In these cases, approximately 3.5 genome equivalents would have to be screened.

In doing partial digests of DNA from a variety of species, it has been observed that a variable proportion of the DNA is resistant to digestion with *Sau*3A and therefore is not included in the library. To allow for this excluded DNA, the following formula can be used to determine the included genome equivalents in the library.

- # of genomes represented = (# of recombinant bp recovered)/(# of recombinant bp/genome) × (% digestible DNA)